

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2004-234555
(P2004-234555A)

(43) 公開日 平成16年8月19日(2004.8.19)

(51) Int. Cl.⁷
G06F 13/00

F I
G06F 13/00 301 J

テーマコード(参考)
5B083

審査請求 未請求 請求項の数 18 O L (全 31 頁)

(21) 出願番号 特願2003-25071 (P2003-25071)
(22) 出願日 平成15年1月31日(2003.1.31)

(71) 出願人 000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目6番地
(74) 代理人 100071283
弁理士 一色 健輔
(74) 代理人 100084906
弁理士 原島 典孝
(74) 代理人 100098523
弁理士 黒川 恵
(74) 代理人 100112748
弁理士 吉田 浩二
(74) 代理人 100110009
弁理士 青木 康

最終頁に続く

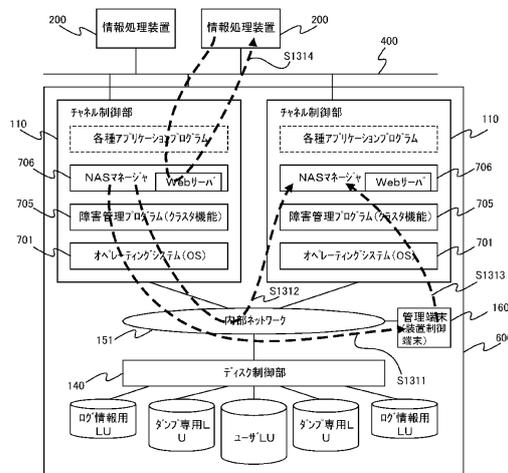
(54) 【発明の名称】 ストレージシステムの制御方法、ストレージシステム、及びプログラム

(57) 【要約】 (修正有)

【課題】 障害の発生を確実に検知し、これを迅速にユーザやオペレータ等に通知して障害原因の究明や復旧などの作業を円滑に進められるようなストレージシステムを提供する。

【解決手段】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有する複数のチャンネル制御部と、I/Oプロセッサから送信されるI/O要求に応じて記憶デバイスに対するデータ入出力を実行するディスク制御部と、チャンネル制御部および前記ディスク制御部と通信可能に接続されるコンピュータとを含んで構成されるストレージシステムに、チャンネル制御部が、チャンネル制御部における障害の発生を検知した場合にその旨を前記コンピュータに通知するようにする。

【選択図】 図13



【特許請求の範囲】**【請求項 1】**

情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I/O 要求を出力する I/O プロセッサとが形成された回路基板を有する複数のチャンネル制御部と、

前記 I/O プロセッサから送信される前記 I/O 要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

前記チャンネル制御部および前記ディスク制御部と通信可能に接続されるコンピュータと、を含んで構成されるストレージシステムの制御方法であって、

前記チャンネル制御部が、前記チャンネル制御部における障害の発生を検知した場合にその旨を前記コンピュータに通知すること、

を特徴とするストレージシステムの制御方法。

10

【請求項 2】

請求項 1 に記載のストレージシステムの制御方法において、

前記チャンネル制御部は、前記チャンネル制御部で実行されるオペレーティングシステムがダンプ情報を出力したことにより前記チャンネル制御部に障害が発生したことを検知すること

、
を特徴とするストレージシステムの制御方法。

20

【請求項 3】

情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I/O 要求を出力する I/O プロセッサとが形成された回路基板を有する複数のチャンネル制御部と、

前記 I/O プロセッサから送信される前記 I/O 要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

を含んで構成されるストレージシステムの制御方法であって、

前記チャンネル制御部が、前記チャンネル制御部の障害を検知した場合にその旨を他のチャンネル制御部に通知すること、

を特徴とするストレージシステムの制御方法。

30

【請求項 4】

情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I/O 要求を出力する I/O プロセッサとが形成された回路基板を有する複数のチャンネル制御部と、

前記 I/O プロセッサから送信される前記 I/O 要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

前記チャンネル制御部および前記ディスク制御部と通信可能に接続されるコンピュータと、を含んで構成されるストレージシステムの制御方法であって、

前記チャンネル制御部が、前記チャンネル制御部の障害を検知した場合にその旨を前記コンピュータに通知し、

前記コンピュータが、前記通知を受信した場合に、前記障害を検知した旨を前記障害を検知した前記チャンネル制御部とは異なる他のチャンネル制御部に通知すること、

を特徴とするストレージシステムの制御方法。

40

【請求項 5】

請求項 4 に記載のストレージシステムの制御方法において、

前記コンピュータは特定の前記チャンネル制御部にのみ前記障害を検知した旨を通知すること、

を特徴とするストレージシステムの制御方法。

【請求項 6】

50

情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有する複数のチャンネル制御部と、

前記I/Oプロセッサから送信される前記I/O要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

前記チャンネル制御部および前記ディスク制御部と通信可能に接続されるコンピュータと、を含んで構成されるストレージシステムの制御方法であって、

前記ディスク制御部が、前記チャンネル制御部から出力されたダンプ情報が前記記憶デバイスに書き込まれることを検知した場合に、前記チャンネル制御部において障害が発生した旨を前記コンピュータに通知すること、

を特徴とするストレージシステムの制御方法。

10

【請求項7】

情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有する複数のチャンネル制御部と、

前記I/Oプロセッサから送信される前記I/O要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

を含んで構成されるストレージシステムの制御方法であって、

前記ディスク制御部が、前記チャンネル制御部から出力されたダンプ情報が前記記憶デバイスに書き込まれることを検知した場合に、前記チャンネル制御部において障害が発生した旨を前記チャンネル制御部とは異なる他の前記チャンネル制御部に通知すること、

を特徴とするストレージシステムの制御方法。

20

【請求項8】

情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有する複数のチャンネル制御部と、

前記I/Oプロセッサから送信される前記I/O要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

前記チャンネル制御部および前記ディスク制御部と通信可能に接続されるコンピュータと、を含んで構成されるストレージシステムの制御方法であって、

前記ディスク制御部が、前記チャンネル制御部から出力されたダンプ情報が前記記憶デバイスに書き込まれることを検知した場合に、前記チャンネル制御部において障害が発生した旨を前記コンピュータに通知し、

前記コンピュータが、前記通知を受信した場合に、前記障害を検知した旨を前記障害を検知した前記チャンネル制御部とは異なる他のチャンネル制御部に通知すること、

を特徴とするストレージシステムの制御方法。

30

【請求項9】

請求項8に記載のストレージシステムの制御方法において、前記コンピュータは特定の前記チャンネル制御部にのみ前記障害を検知した旨を通知すること、

を特徴とするストレージシステムの制御方法。

40

【請求項10】

請求項6乃至9のいずれかに記載のストレージシステムの制御方法において、

前記ダンプ情報は、前記記憶デバイスの記憶領域に確保された専用の記憶領域に書き込まれること、

を特徴とするストレージシステムの制御方法。

【請求項11】

請求項1乃至9のいずれかに記載のストレージシステムの制御方法において、

50

前記チャンネル制御部は、前記チャンネル制御部において障害が発生した旨の通知を前記コンピュータもしくは他の前記チャンネル制御部から受信した場合に、その旨を前記情報処理装置に対して通知すること、
を特徴とするストレージシステムの制御方法。

【請求項 1 2】

請求項 1 乃至 9 のいずれかに記載のストレージシステムの制御方法において、
前記チャンネル制御部は、前記チャンネル制御部において障害が発生した旨の通知を受信した場合に、障害が発生している前記チャンネル制御部により書き込まれたダンプ情報の前記記憶デバイス上の所在を前記通知に基づいて特定し、
前記チャンネル制御部は、前記特定された前記所在に記憶されている前記ダンプ情報を取得してそのダンプ情報を前記情報処理装置に通知すること、
を特徴とするストレージシステムの制御方法。

10

【請求項 1 3】

請求項 1 2 に記載のストレージシステムの制御方法において、
前記チャンネル制御部は、前記記憶デバイスに記憶されている前記ダンプ情報を、前記情報処理装置を操作する人に参照させる形式に加工した形で前記情報処理装置に通知すること、
を特徴とするストレージシステムの制御方法。

【請求項 1 4】

請求項 1 乃至 9 のいずれかに記載のストレージシステムの制御方法において、
前記障害は、前記チャンネル制御部で実行されるオペレーティングシステムもしくはアプリケーションプログラムのいずれかの障害に起因して生じたものであること、
を特徴とするストレージシステムの制御方法。

20

【請求項 1 5】

情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I/O 要求を出力する I/O プロセッサとが形成された回路基板を有する複数のチャンネル制御部と、
前記 I/O プロセッサから送信される前記 I/O 要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、
前記チャンネル制御部および前記ディスク制御部と通信可能に接続されるコンピュータとを含み、
前記チャンネル制御部が、前記チャンネル制御部における障害の発生を検知した場合にその旨を前記コンピュータに通知する手段
を備えることを特徴とするストレージシステム。

30

【請求項 1 6】

情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I/O 要求を出力する I/O プロセッサとが形成された回路基板を有する複数のチャンネル制御部と、
前記 I/O プロセッサから送信される前記 I/O 要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、
前記チャンネル制御部および前記ディスク制御部と通信可能に接続されるコンピュータとを含み、
前記ディスク制御部が、前記チャンネル制御部から出力されたダンプ情報が前記記憶デバイスに書き込まれることを検知した場合に、前記チャンネル制御部において障害が発生した旨を前記コンピュータに通知する手段
を備えることを特徴とするストレージシステムの制御方法。

40

【請求項 1 7】

情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて

50

受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有する複数のチャンネル制御部と、
前記I/Oプロセッサから送信される前記I/O要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、
前記チャンネル制御部および前記ディスク制御部と通信可能に接続されるコンピュータとを含んで構成されるストレージシステムにおける前記チャンネル制御部に、前記チャンネル制御部における障害の発生を検知した場合にその旨を前記コンピュータに通知する機能を実現させるためのプログラム。

【請求項18】

情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有する複数のチャンネル制御部と、
前記I/Oプロセッサから送信される前記I/O要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、
前記チャンネル制御部および前記ディスク制御部と通信可能に接続されるコンピュータとを含んで構成されるストレージシステムにおける前記ディスク制御部に、前記チャンネル制御部から出力されたダンプ情報が前記記憶デバイスに書き込まれることを検知した場合に、前記チャンネル制御部において障害が発生した旨を前記コンピュータに通知する機能を実現させるためのプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、ストレージシステムの制御方法、ストレージシステム、及びプログラムに関する。

【0002】

【従来の技術】

近年コンピュータシステムで取り扱われるデータ量が急激に増加している。このようなデータを管理するためのストレージシステムとして、最近ではミッドレンジクラスやエンタープライズクラスと呼ばれるような、巨大な記憶資源を提供するRAID (Redundant Arrays of Inexpensive Disks) 方式で管理された大規模なストレージシステムが注目されている。また、かかる膨大なデータを効率よく利用し管理するために、ディスクアレイ装置等のストレージシステムと情報処理装置とを専用のネットワーク (Storage Area Network、以下SANと記す) で接続し、ストレージシステムへの高速かつ大量なアクセスを実現する技術が開発されている。

【0003】

一方、ストレージシステムと情報処理装置とをTCP/IP (Transmission Control Protocol / Internet Protocol) プロトコル等を用いたネットワークで相互に接続し、情報処理装置からのファイルレベルでのアクセスを実現するNAS (Network Attached Storage) と呼ばれるストレージシステムが開発されている。

【0004】

【特許文献1】

特開平9-50386号公報

【0005】

【発明が解決しようとする課題】

ところで、膨大な量のデータを取り扱うストレージシステムに何らかの障害が発生し、これによりサービスが停止した場合には、ユーザに多大な損失を生じさせることになる。こ

10

20

30

40

50

のためストレージシステムに対しては高可用が求められる。そしてこのような高可用の要請を満たすためには、障害の発生を確実に検知しこれを迅速にユーザやオペレータ等に通知して障害原因の究明や復旧などの作業を円滑に進められるよう、ストレージシステムには効率的な障害管理の仕組みが求められる。

本発明は上記課題を鑑みてなされたものであり、ストレージシステムの制御方法、ストレージシステム、及びプログラムを提供することを主たる目的とする。

【0006】

【課題を解決するための手段】

上記目的を達成するための本発明の主たる発明は、

情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有する複数のチャンネル制御部と、

前記I/Oプロセッサから送信される前記I/O要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

前記チャンネル制御部および前記ディスク制御部と通信可能に接続されるコンピュータと、を含んで構成されるストレージシステムの制御方法であって、

前記チャンネル制御部が、前記チャンネル制御部における障害の発生を検知した場合にその旨を前記コンピュータに通知することとする。

【0007】

なお、前記情報処理装置とは、前記構成の前記ストレージシステムにLANやSANを介してアクセスする、例えば、パーソナルコンピュータやメインフレームコンピュータである。ファイルアクセス処理部の機能はCPU上で実行されるオペレーティングシステムおよびこのオペレーティングシステム上で動作する例えばNFS(Network File System)等のソフトウェアによって提供される。記憶デバイスは例えばハードディスク装置などのディスクドライブである。I/Oプロセッサは例えばファイルアクセス処理部のハードウェア要素である前記CPUとは独立したIC(Integrated Circuit)をハードウェア要素とし、ファイルアクセス処理部とディスク制御部との間の通信を制御する。ディスク制御部は、記憶デバイスに対してデータの書き込みや読み出しを行う。

その他、本願が開示する課題、及びその解決方法は、発明の実施の形態の欄、及び図面により明らかにされる。

【0008】

【発明の実施の形態】

===全体構成例===

ストレージシステム600は、記憶デバイス制御装置100と記憶デバイス(記憶装置)300とを備えている。記憶デバイス制御装置100は、情報処理装置200から受信したコマンドに従って記憶デバイス300に対する制御を行う。例えば情報処理装置200からデータの入出力要求を受信して、記憶デバイス300に記憶されているデータの入出力のための処理を行う。データは、記憶デバイス300が備えるディスクドライブにより提供される物理的な記憶領域上に論理的に設定される記憶領域である論理ボリューム(Logical Unit)(以下、LUと記す)に記憶されている。また記憶デバイス制御装置100は、情報処理装置200との間で、ストレージシステム600を管理するための各種コマンドの授受も行う。

【0009】

情報処理装置200はCPU(Central Processing Unit)やメモリを備えたコンピュータである。情報処理装置200が備えるCPUにより各種プログラムが実行されることにより様々な機能が実現される。情報処理装置200は、例えばパーソナルコンピュータやワークステーションであることもあるし、メインフレームコンピュータであることもある。

10

20

30

40

50

【0010】

図1において、情報処理装置200は、外部のネットワークである外部LAN(Local Area Network)400を介して記憶デバイス制御装置100と接続されている。外部LAN400は、インターネットとすることもできるし、専用のネットワークとすることもできる。外部LAN400を介して行われる情報処理装置200と記憶デバイス制御装置100との間の通信は、例えばTCP/IPプロトコルに従って行われる。情報処理装置200からは、ストレージシステム600に対して、ファイル名指定によるデータアクセス要求(以下、ファイルアクセス要求とも記す)が送信される。

【0011】

記憶デバイス制御装置100は、チャンネル制御部110を備える。以下、このチャンネル制御部110のことをCHNとも記す。記憶デバイス制御装置100は、チャンネル制御部110により外部LAN400を介して情報処理装置200との間で通信を行う。チャンネル制御部110は、情報処理装置200からのファイルアクセス要求を個々に受け付ける。すなわち、チャンネル制御部110には、個々に外部LAN400上のネットワークアドレス(例えば、IPアドレス)が割り当てられていてそれぞれが個別にNASとして振る舞い、個々のNASがあたかも独立したNASが存在しているかのようにNASのサービスを情報処理装置200に提供することができる。以下、チャンネル制御部110をCHNと記す。このように1台のストレージシステム600に個別にNASとしてのサービスを提供するチャンネル制御部110を備えるように構成したことで、従来、独立したコンピュータで個別に運用されていたNASサーバが一台のストレージシステム600に集約されて運用される。そして、この構成によってストレージシステム600の統括的な管理が可能となり、各種設定・制御や障害管理、バージョン管理といった保守業務の効率化が図られる。

【0012】

なお、本実施の形態に係る記憶デバイス制御装置100のチャンネル制御部110の機能は、後述するように、一体的にユニット化された回路基板上に形成されたハードウェア及びこのハードウェアにより実行されるオペレーティングシステム(以下、OSと記す)やこのOS上で動作するアプリケーションプログラムなどのソフトウェアにより実現される。このように本実施例のストレージシステム600では、従来ハードウェアの一部として実装されてきた機能が主としてソフトウェアによって実現されている。このため、本実施例のストレージシステム600では柔軟性に富んだシステム運用が可能となり、多様で変化の激しいユーザニーズに対応したきめ細かなサービスを提供することが可能となる。

【0013】

=== 記憶デバイス ===

記憶デバイス300は多数のディスクドライブ(物理ディスク)を備えており情報処理装置200に対して記憶領域を提供する。データは、ディスクドライブにより提供される物理的な記憶領域上に論理的に設定される記憶領域であるLUに記憶されている。ディスクドライブとしては、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置等様々なものを用いることができる。

【0014】

なお、記憶デバイス300は例えば複数のディスクドライブによりディスクアレイを構成するようにすることもできる。この場合、情報処理装置200に対して提供される記憶領域は、RAIDにより管理された複数のディスクドライブにより提供されるようにすることもできる。

【0015】

記憶デバイス制御装置100と記憶デバイス300との間は図1のように直接に接続される形態とすることもできるし、ネットワークを介して接続するようにすることもできる。さらに記憶デバイス300は記憶デバイス制御装置100と一体として構成されることもできる。

【0016】

10

20

30

40

50

記憶デバイス 300 に設定される LU には、情報処理装置 200 からアクセス可能なユーザ LU や、チャンネル制御部 110 の制御のために使用されるシステム LU 等がある。システム LU には CHN 110 で実行されるオペレーティングシステムも格納される。また各 LU にはチャンネル制御部 110 が対応付けられている。これによりチャンネル制御部 110 毎にアクセス可能な LU が割り当てられている。また上記対応付けは、複数のチャンネル制御部 110 で一つの LU を共有するようにすることもできる。なお以下において、ユーザ LU やシステム LU をユーザディスク、システムディスク等とも記す。また、複数のチャンネル制御部 110 で共有される LU を共有 LU あるいは共有ディスクと記す。

【0017】

=== 記憶デバイス制御装置 ===

記憶デバイス制御装置 100 は、チャンネル制御部 110、共有メモリ 120、キャッシュメモリ 130、ディスク制御部 140、管理端末 160、接続部 150 を備える。

【0018】

チャンネル制御部 110 は情報処理装置 200 との間で通信を行うための通信インタフェースを備え、情報処理装置 200 との間でデータ入出力コマンド等を授受する機能を備える。例えば CHN 110 は情報処理装置 200 からのファイルアクセス要求を受け付ける。そしてファイルの記憶アドレスやデータ長等を求めて、ファイルアクセス要求に対応する I/O 要求を出力することにより、記憶デバイス 300 へのアクセスを行う。これによりストレージシステム 600 は NAS としてのサービスを情報処理装置 200 に提供することができる。なお I/O 要求にはデータの先頭アドレス、データ長、読み出し又は書き込み等のアクセスの種別が含まれている。またデータの書き込みの場合には I/O 要求には書き込みデータが含まれているようにすることもできる。I/O 要求の出力は、後述する I/O プロセッサ 119 により行われる。

【0019】

各チャンネル制御部 110 は管理端末（装置制御端末）160 と共にストレージシステム 600 内部のネットワークである内部 LAN 151 で接続されている。これによりチャンネル制御部 110 に実行させるマイクロプログラム等を管理端末 160 から送信しインストールすることが可能となっている。チャンネル制御部 110 の構成については後述する。

【0020】

接続部 150 はチャンネル制御部 110、共有メモリ 120、キャッシュメモリ 130、ディスク制御部 140 を相互に接続する。チャンネル制御部 110、共有メモリ 120、キャッシュメモリ 130、ディスク制御部 140 間でのデータやコマンドの授受は接続部 150 を介することにより行われる。接続部 150 は例えば高速スイッチングによりデータ伝送を行う超高速クロスバススイッチなどの高速バスである。チャンネル制御部 110 同士が高速バスで接続されていることで、個々のコンピュータ上で動作する NAS サーバが LAN を通じて接続する従来の構成に比べてチャンネル制御部 110 間の通信パフォーマンスが大幅に向上する。またこれにより高速なファイル共有機能や高速フェイルオーバーなども可能となる。

【0021】

共有メモリ 120 及びキャッシュメモリ 130 は、チャンネル制御部 110、ディスク制御部 140 により共有される記憶メモリである。共有メモリ 120 は主に制御情報やコマンド等を記憶するために利用されるのに対し、キャッシュメモリ 130 は主にデータを記憶するために利用される。

【0022】

例えば、あるチャンネル制御部 110 が情報処理装置 200 から受信したデータ入出力コマンドが書き込みコマンドであった場合には、当該チャンネル制御部 110 は書き込みコマンドを共有メモリ 120 に書き込むと共に、情報処理装置 200 から受信した書き込みデータをキャッシュメモリ 130 に書き込む。一方、ディスク制御部 140 は共有メモリ 120 を監視しており、共有メモリ 120 に書き込みコマンドが書き込まれたことを検出すると、当該コマンドに従ってキャッシュメモリ 130 から書き込みデータを読み出して記憶

10

20

30

40

50

デバイス 300 に書き込む。

【0023】

またあるチャンネル制御部 110 が情報処理装置 200 から受信したデータ入出力コマンドが読み出しコマンドであった場合には、当該チャンネル制御部 110 は読み出しコマンドを共有メモリ 120 に書き込むと共に、読み出し対象となるデータがキャッシュメモリ 130 に存在するかどうかを調べる。ここでキャッシュメモリ 130 に存在すれば、そのデータを情報処理装置 200 に送信する。一方、読みだし対象となるデータがキャッシュメモリ 130 に存在しない場合には、共有メモリ 120 を監視することにより読み出しコマンドが共有メモリ 120 に書き込まれたことを検出したディスク制御部 140 が、記憶デバイス 300 から読みだし対象となるデータを読み出してこれをキャッシュメモリ 130 に書き込むと共に、その旨を共有メモリ 120 に書き込む。そして、チャンネル制御部 110 は共有メモリ 120 の監視により読みだし対象となるデータがキャッシュメモリ 130 に書き込まれたことを検出すると、そのデータを情報処理装置 200 に送信する。

10

【0024】

なお、このようにチャンネル制御部 110 からディスク制御部 140 に対するデータの書き込みや読み出しの指示を共有メモリ 120 を介在させて間接に行う構成の他、例えばチャンネル制御部 110 からディスク制御部 140 に対してデータの書き込みや読み出しの指示を共有メモリ 120 を介さずに直接に行う構成とすることもできる。

【0025】

ディスク制御部 140 は記憶デバイス 300 の制御を行う。例えば上述のように、チャンネル制御部 110 が情報処理装置 200 から受信したデータ書き込みコマンドに従って記憶デバイス 300 へデータの書き込みを行う。また、チャンネル制御部 110 により送信された論理アドレス指定による LU へのデータアクセス要求を、物理アドレス指定による物理ディスクへのデータアクセス要求に変換する。記憶デバイス 300 における物理ディスクが RAID により管理されている場合には、RAID 構成（例えば、RAID0, 1, 5）に従ったデータのアクセスを行う。またディスク制御部 140 は、記憶デバイス 300 に記憶されたデータの複製管理の制御やバックアップ制御を行う。さらにディスク制御部 140 は、災害発生時のデータ消失防止（ディザスタリカバリ）などを目的としてプライマリサイトのストレージシステム 600 のデータの複製をセカンダリサイトに設置された他のストレージシステムにも記憶する制御（レプリケーション機能、又はリモートコピー機能）なども行う。

20

30

【0026】

各ディスク制御部 140 は管理端末 160 と共に内部 LAN 151 で接続されており、相互に通信を行うことが可能である。これにより、ディスク制御部 140 に実行させるマイクロプログラム等を管理端末 160 から送信しインストールすることが可能となっている。ディスク制御部 140 の構成については後述する。

【0027】

=== 管理端末 ===

管理端末 160（装置制御端末）はストレージシステム 600 の保守・管理、ストレージシステム 600 におけるハードウェアやソフトウェアの動作状態や実行状態や障害発生の監視等をするためのコンピュータである。管理端末 160 を操作することにより、例えば記憶デバイス 300 内の物理ディスク構成の設定や、LU の設定、チャンネル制御部 110 において実行されるマイクロプログラムのインストール等を行うことができる。ここで、記憶デバイス 300 内の物理ディスク構成の設定としては、例えば物理ディスクの増設や減設、RAID 構成の変更（例えば RAID1 から RAID5 への変更等）等を行うことができる。さらに管理端末 160 からは、ストレージシステム 600 の動作状態の確認や故障部位の特定、チャンネル制御部 110 で実行されるオペレーティングシステムのインストール等の作業を行うこともできる。また管理端末 160 は LAN や電話回線等で外部保守センタと接続されており、管理端末 160 を利用してストレージシステム 600 の障害監視を行ったり、障害が発生した場合に迅速に対応することも可能である。障害の発生は

40

50

例えばチャンネル制御部で動作するOS701やアプリケーションプログラム、ドライバソフトウェアなどから通知される。この通知はHTTPプロトコルやSNMP(Simple Network Management Protocol)、電子メールなどにより行われる。これらの設定や制御は、管理端末160で動作するWebサーバが提供するWebページをユーザインタフェースとしてオペレータなどにより行われる。オペレータ等は、管理端末160を操作して障害監視する対象や内容の設定、障害通知先の設定などを行うこともできる。

【0028】

管理端末160は記憶デバイス制御装置100に内蔵されている形態とすることもできるし、外付けされている形態とすることもできる。また管理端末160は、記憶デバイス制御装置100及び記憶デバイス300の保守・管理を専用に行うコンピュータとすることもできるし、汎用のコンピュータに保守・管理機能を持たせたものとすることもできる。

10

【0029】

管理端末160の構成を示すブロック図を図2に示す。

管理端末160は、CPU161、メモリ162、ポート163、記録媒体読取装置164、入力装置165、出力装置166、記憶装置168を備える。CPU161は管理端末160の全体の制御を司るもので、メモリ162に格納されたプログラム162cを実行することにより上記Webサーバとしての機能等を実現する。メモリ162には、物理ディスク管理テーブル162aとLU管理テーブル162bとプログラム162cとが記憶されている。物理ディスク管理テーブル162aは、記憶デバイス300に備えられる物理ディスク(ディスクドライブ)を管理するためのテーブルである。物理ディスク管理テーブル162aを図3に示す。図3においては、記憶デバイス300が備える多数の物理ディスクのうち、ディスク番号#001乃至#006までが示されている。それぞれの物理ディスクに対して、容量、RAID構成、使用状況が示されている。

20

【0030】

LU管理テーブル162bは、上記物理ディスク上に論理的に設定されるLUを管理するためのテーブルである。LU管理テーブル162bを図4に示す。図4においては、記憶デバイス300上に設定される多数のLUのうち、LU番号#1乃至#3までが示されている。それぞれのLUに対して、物理ディスク番号、容量、RAID構成が示されている。

30

【0031】

記録媒体読取装置164は、記録媒体167に記録されているプログラムやデータを読み取るための装置である。読み取られたプログラムやデータはメモリ162や記憶装置168に格納される。従って、例えば記録媒体167に記録されたプログラム162cを、記録媒体読取装置164を用いて上記記録媒体167から読み取って、メモリ162や記憶装置168に格納するようすることができる。なお、記録媒体167に記録されるプログラムとは、例えば、管理端末160を機能させるためのプログラムの他、チャンネル制御部110のOSやアプリケーションプログラムのインストールやバージョンアップのためのプログラムである。記録媒体167としてはフレキシブルディスクやCD-ROM、DVD-ROM、DVD-RAM、半導体メモリ等を用いることができる。なお、上記プログラム162cは管理端末160を動作させるためのプログラムとすることもできる他、チャンネル制御部110やディスク制御部140にOS701やアプリケーションプログラムをインストールするためのプログラムや、バージョンアップするためのプログラムとすることもできる。記録媒体読取装置164は管理端末160に内蔵されている形態とすることもできるし、外付けされている形態とすることもできる。記憶装置168は、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置等である。入力装置165はオペレータ等による管理端末160へのデータ入力等のために用いられる。入力装置165としては例えばキーボードやマウス等が用いられる。出力装置166は情報を外部に出力するための装置である。出力装置166としては例えばディスプレイやプリンタ等が用いられる。ポート163は内部LAN151に接続されており、これにより管理端末

40

50

160はチャンネル制御部110やディスク制御部140等と通信を行うことができる。またポート163は、外部ネットワークである外部LAN400に接続するようにすることもできるし、電話回線に接続するようにすることもできる。

【0032】

=== 外観図 ===

次に、本実施の形態に係るストレージシステム600の外観構成を図5に示す。また、記憶デバイス制御装置100の外観構成を図6に示す。

図5に示すように、本実施の形態に係るストレージシステム600は記憶デバイス制御装置100及び記憶デバイス300がそれぞれの筐体に納められた形態をしている。記憶デバイス制御装置100の筐体の両側に記憶デバイス300の筐体が配置されている。

10

【0033】

記憶デバイス制御装置100は、正面中央部に管理端末160が備えられている。管理端末160はカバーで覆われており、図6に示すようにカバーを開けることにより管理端末160を使用することができる。なお図6に示した管理端末160はいわゆるノート型パーソナルコンピュータの形態をしているが、どのような形態とすることも可能である。

【0034】

管理端末160の下部には、チャンネル制御部110のボードを装着するためのスロットが設けられている。チャンネル制御部110のボードとは、チャンネル制御部110の回路基板が形成されたユニットであり、スロットへの装着単位である。本実施の形態に係るストレージシステム600においては、スロットは8つあり、図5及び図6には8つのスロットにチャンネル制御部110のハードウェアを構成する回路が形成されたボードが装着された状態が示されている。各スロットにはチャンネル制御部110のボードを装着するためのガイドレールが設けられている。ガイドレールに沿ってチャンネル制御部110のボードをスロットに挿入することにより、チャンネル制御部110のボードを記憶デバイス制御装置100に装着することができる。また各スロットに装着されたチャンネル制御部110のボードは、ガイドレールに沿って手前方向に引き抜くことにより取り外すことができる。また各スロットの奥手方向正面部には、各チャンネル制御部110のボードを記憶デバイス制御装置100と電気的に接続するためのコネクタが設けられている。

20

【0035】

なお、スロットには以上に説明したNASとして機能するタイプのチャンネル制御部110以外にも、SAN(Storage Area Network)に接続する機能を備えるタイプや、FICON(Fibre Connection)(登録商標)やESCON(Enterprise System Connection)(登録商標)等のメインフレーム系のプロトコルに従って通信を行う機能を備えるタイプのチャンネル制御部110が装着されることもある。またチャンネル制御部110のボードを装着しないスロットを設けるようにすることもできる。

30

【0036】

各スロットのチャンネル制御部110は、同種の複数のチャンネル制御部110でクラスタを構成する。例えば2枚のCHN110をペアとしてクラスタを構成することができる。クラスタを構成することにより、クラスタ内のあるチャンネル制御部110に障害が発生した場合でも、障害が発生したチャンネル制御部110がそれまで行っていた処理をクラスタ内の他のチャンネル制御部110に引き継ぐようにすることができる(フェイルオーバー制御)。2枚のCHN110でクラスタを構成している様子を示す図を図12に示すが、詳細は後述する。

40

【0037】

なお、記憶デバイス制御装置100は信頼性向上のため電源供給が2系統化されており、チャンネル制御部110のボードが装着される上記8つのスロットは電源系統毎に4つずつに分けられている。そこでクラスタを構成する場合には、両方の電源系統のチャンネル制御部110のボードを含むようにする。これにより、片方の電源系統に障害が発生し電力の供給が停止しても、同一クラスタを構成する他方の電源系統に属するチャンネル制御部11

50

0のボードへの電源供給は継続されるため、当該チャンネル制御部110に処理を引き継ぐ(フェイルオーバー)ことができる。

【0038】

なお、上述したように、チャンネル制御部110は上記各スロットに装着可能なボードとして提供されるが、上記一つのボードは一体形成された複数枚数の回路基板から構成されているようにすることもできる。

ディスク制御部140や共有メモリ120等の、記憶デバイス制御装置100を構成する他の装置については図5及び図6には示されていないが、記憶デバイス制御装置100の背面側等に装着されている。

また記憶デバイス制御装置100には、チャンネル制御部110のボード等から発生する熱を放出するためのファン170が設けられている。ファン170は記憶デバイス制御装置100の上面部に設けられる他、チャンネル制御部110用スロットの上部にも設けられている。

【0039】

ところで、筐体に収容されて構成される記憶デバイス制御装置100および記憶デバイス300としては、例えばSAN対応として製品化されている従来構成の装置を利用することができる。特に上記のようにCHN110のボードのコネクタ形状を従来構成の筐体に設けられているスロットにそのまま装着できる形状とすることで従来構成の装置をより簡単に利用することができる。つまり本実施例のストレージシステム600は、既存の製品を利用することで容易に構築することができる。

【0040】

===チャンネル制御部===

本実施の形態に係るストレージシステム600は、上述の通りCHN110により情報処理装置200からのファイルアクセス要求を受け付け、NASとしてのサービスを情報処理装置200に提供する。

CHN110のハードウェア構成を図7に示す。この図に示すようにCHN110のハードウェアは一体的にユニット化されたボードで構成される。以下、このユニットのことをNASボードとも記す。NASボードは一枚もしくは複数枚の回路基板を含んで構成される。より具体的には、NASボードは、ネットワークインタフェース部111、CPU112、メモリ113、入出力制御部114、I/O(Input/Output)プロセッサ119、NVRAM(Non Volatile RAM)115、ボード接続用コネクタ116、通信コネクタ117を備え、これらが一体的にユニット化された回路基板上に形成されて構成されている。

【0041】

ネットワークインタフェース部111は、情報処理装置200との間で通信を行うための通信インタフェースを備えている。CHN110の場合は、例えばTCP/IPプロトコルに従って情報処理装置200から送信されたファイルアクセス要求を受信する。通信コネクタ117は情報処理装置200と通信を行うためのコネクタである。CHN110の場合は外部LAN400に接続可能なコネクタであり、例えばイーサネット(登録商標)に対応している。

CPU112は、CHN110をNASボードとして機能させるための制御を司る。

【0042】

メモリ113には様々なプログラムやデータが記憶される。例えば図8に示すメタデータ730やロックテーブル720、また図10に示すNASマネージャ706等の各種プログラムが記憶される。

【0043】

図9にメタデータ730の一例を示す。メタデータ730はファイルシステムプログラム703により実現されるファイルシステムが管理しているファイルに対応させて生成される情報である。メタデータ730には例えばファイルのデータが記憶されているLU上のアドレスやデータサイズなど、ファイルの保管場所を特定するための情報が含まれる。メ

タデータ730にはファイルの容量、所有者、更新時刻等の情報が含まれることもある。また、メタデータ730はファイルだけでなくディレクトリに対応させて生成されることもある。メタデータ730は記憶デバイス300上の各LUにも記憶されている。

【0044】

入出力制御部114は、ディスク制御部140やキャッシュメモリ130、共有メモリ120、管理端末160との間でデータやコマンドの授受を行う。入出力制御部114はI/Oプロセッサ119やNVRAM115を備えている。I/Oプロセッサ119は例えば1チップのマイコンで構成される。I/Oプロセッサ119は上記データやコマンドの授受を制御し、CPU112とディスク制御部140との間の通信を中継する。NVRAM115はI/Oプロセッサ119の制御を司るプログラムを格納する不揮発性メモリである。NVRAM115に記憶されるプログラムの内容は、管理端末160や、後述するNASマネージャ706からの指示により書き込みや書き換えを行うことができる。

10

【0045】

ディスク制御部140のハードウェア構成を示す図を図10に示す。

ディスク制御部140は、一体的にユニット化されたボードとして形成されている。ディスク制御部140のボードは、インタフェース部141、メモリ143、CPU142、NVRAM144、ボード接続用コネクタ145を備え、これらが一体的にユニット化された回路基板上に形成されている。

インタフェース部141は、接続部150を介してチャンネル制御部110等との間で通信を行うための通信インタフェースや、記憶デバイス300との間で通信を行うための通信インタフェースを備えている。

20

【0046】

CPU142は、ディスク制御部140全体の制御を司ると共に、チャンネル制御部110や記憶デバイス300、管理端末160との間の通信を行う。メモリ143やNVRAM144に格納された各種プログラムを実行することにより本実施の形態に係るディスク制御部140の機能が実現される。ディスク制御部140により実現される機能としては、記憶デバイス300の制御やRAID制御、記憶デバイス300に記憶されたデータの複製管理やバックアップ制御、リモートコピー制御等である。

【0047】

NVRAM144はCPU142の制御を司るプログラムを格納する不揮発性メモリである。NVRAM144に記憶されるプログラムの内容は、管理端末160や、NASマネージャ706からの指示により書き込みや書き換えを行うことができる。

30

またディスク制御部140のボードはボード接続用コネクタ145を備えている。ボード接続用コネクタ145が記憶デバイス制御装置100側のコネクタと嵌合することにより、ディスク制御部140のボードは記憶デバイス制御装置100と電氣的に接続される。

【0048】

===ソフトウェア構成===

次に、本実施の形態に係るストレージシステム600におけるソフトウェア構成図を図11に示す。

オペレーティングシステム701は例えばUNIX(登録商標)である。オペレーティングシステム701上では、RAIDマネージャ708、ボリュームマネージャ707、SVPマネージャ709、ファイルシステムプログラム703、ネットワーク制御部702、障害管理プログラム705、NASマネージャ706などのソフトウェアが動作する。

40

【0049】

オペレーティングシステム701上で動作するRAIDマネージャ708は、RAID制御部740に対するパラメータの設定やRAID制御部740を制御する機能を提供する。RAIDマネージャ708はオペレーティングシステム701やオペレーティングシステム701上で動作する他のアプリケーション、もしくは管理端末160からパラメータや制御指示情報を受け付けて、受け付けたパラメータのRAID制御部740への設定や、RAID制御部指示情報に対応する制御コマンドの送信を行う。

50

【 0 0 5 0 】

ここで設定されるパラメータとしては、例えば、RAIDグループを構成する記憶デバイス（物理ディスク）を定義（RAIDグループの構成情報、ストライプサイズの指定など）するためのパラメータ、RAIDレベル（例えば0, 1, 5）を設定するためのパラメータなどがある。また、RAIDマネージャ708がRAID制御部740に送信する制御コマンドとしてはRAIDの構成・削除・容量変更を指示するコマンド、各RAIDグループの構成情報を要求するコマンドなどがある。

【 0 0 5 1 】

ボリュームマネージャ707は、RAID制御部740によって提供されるLUをさらに仮想化した仮想化論理ボリュームをファイルシステムプログラム703に提供する。1つの仮想化論理ボリュームは1以上の論理ボリュームによって構成される。なお、OS701およびこの上で実行される各種のアプリケーションプログラムは、LUに直接アクセスすることもできるし、仮想化論理ボリュームに対してアクセスすることもできる。

【 0 0 5 2 】

ファイルシステムプログラム703の主な機能は、ネットワーク制御部702が受信したファイルアクセス要求に指定されているファイル名とそのファイル名が格納されているLUもしくは仮想化論理ボリューム上のアドレスとの対応づけを管理することである。例えば、ファイルシステムプログラム703はファイルアクセス要求に指定されているファイル名に対応するLU上のアドレスもしくは仮想化論理ボリューム上のアドレスを特定する。

【 0 0 5 3 】

ネットワーク制御部702は、NFS（Network File System）711とCIFS（Common Interface File System）713の2つのファイルシステムプロトコルを含んで構成される。NFS711は、NFS711が動作するUNIX（登録商標）系の情報処理装置200からのファイルアクセス要求を受け付ける。一方、CIFS713はCIFS713が動作するWindows（登録商標）系の情報処理装置200からのファイルアクセス要求を受け付ける。

【 0 0 5 4 】

NASマネージャ706は、ストレージシステム600について、その動作状態の確認、設定や制御などを行うためのプログラムである。NASマネージャ706は外部LAN400に接続している情報処理装置200に対するWebサーバとしての機能も有し、情報処理装置200からストレージシステム600の設定や制御、ストレージシステム600の動作状態の監視や障害発生の有無の監視を行うための設定Webページを情報処理装置200に提供する。設定Webページはチャンネル制御部110の個々において動作するNASマネージャ706により提供される。NASマネージャ706は、情報処理装置200からのHTTP（HyperText Transport Protocol）リクエストに応じて、設定Webページのデータを情報処理装置200に送信する。情報処理装置200に表示された設定Webページを利用してシステムアドミニストレータなどによりストレージシステム600の設定や制御の指示が行われる。

【 0 0 5 5 】

NASマネージャ706は、設定Webページに対する操作に起因して情報処理装置200から送信される設定や制御に関するデータを受信してそのデータに対応する設定や制御を実行する。これにより、情報処理装置200からストレージシステム600の様々な設定や制御を行うことができる。またNASマネージャ706は設定Webページの設定内容をチャンネル制御部110上で動作するOSやアプリケーションプログラム、ディスク制御部140等に通知する。設定Webページで設定された内容は共有LU310に管理されることもある。

【 0 0 5 6 】

NASマネージャ706の設定Webページを利用して行うことができる内容としては、例えば、LUの管理や設定（容量管理や容量拡張・縮小、ユーザ割り当て等）、上述の複

10

20

30

40

50

製管理やリモートコピー（レプリケーション）等の機能に関する設定や制御（複製元のLUと複製先のLUの設定など）、冗長構成されたCHN110のクラスタの管理（フェイルオーバーさせる相手の対応関係の設定、フェイルオーバー方法など）、OSやOS上で動作するアプリケーションプログラムのバージョン管理、などがある。

【0057】

なお、NASマネージャ706によるストレージシステム600の動作状態の確認、設定や制御には、上述した設定Webを介する方法以外に、クライアント・サーバシステムとすることもできる。その場合、NASマネージャ706はクライアント・サーバシステムのサーバ機能を有し、情報処理装置200のクライアント機能の操作に起因して送信される設定や制御を、上述した設定Webページと同じようにして実施することで、ストレージシステム600の動作状態の確認、設定や制御を行う。

10

【0058】

障害管理プログラム705は、クラスタを構成するチャンネル制御部110間でのフェイルオーバー制御を行うためのプログラムである。

2枚のCHN110でクラスタ180が構成されている様子を示す図を図12に示す。図12では、CHN1（チャンネル制御部1）110とCHN2（チャンネル制御部2）110とでクラスタ180が構成されている場合を示す。

【0059】

上述したように、フェイルオーバー処理はクラスタ180を構成するチャンネル制御部110間で行われる。つまり、例えばCHN1（110）に何らかの障害が発生し処理を継続することができなくなった場合には、CHN1（110）がそれまで行っていた処理はCHN2（110）に引き継がれる。フェイルオーバー処理は、CHN1（110）とCHN2（110）により実行される障害管理プログラム705により実行される。

20

【0060】

CHN1（110）及びCHN2（110）は共に障害管理プログラム705を実行し、共有メモリ120に対して自己の処理が正常に行われていることを書き込むようにする。そして、相手側の上記書き込みの有無を相互に確認するようにする。相手側による書き込みが検出できない場合には、相手側に何らかの障害が発生したと判断し、フェイルオーバー処理を実行する。フェイルオーバー実行時の処理の引き継ぎは、共有LU310を介して行われる。

30

【0061】

フェイルオーバーはこのように自動的に行われることもあるが、オペレータが管理端末160を操作して指示を与えることにより手動で行われることもある。またユーザがNASマネージャ706が提供する設定Webページを利用することにより、情報処理装置200側からの手動による指示により行われることもある。このようにフェイルオーバーを手動で行う目的としては、耐用年数の経過やバージョンアップ、定期診断などのためにチャンネル制御部110のハードウェア（例えばNASボード）を交換する必要が生じた場合などがある。

【0062】

SVPMマネージャ709は、管理端末160からの要求に応じて各種のサービスを管理端末160に提供する。例えば、LUの設定内容やRAIDの設定内容等のストレージシステム600に関する各種設定内容の管理端末160への提供や、管理端末160から入力されたストレージシステム600に関する各種設定の反映等を行う。

40

【0063】

=== 障害管理の仕組み ===

次に本実施例のストレージシステム600における障害管理の仕組みについて図13とともに説明する。なお、この図においては説明に必要なハードウェア及びソフトウェアのみを示している。上述のように各チャンネル制御部110ではOS701及びアプリケーションプログラムが実行されている。この図に示すように、各チャンネル制御部110で実行されているOS701やアプリケーションプログラムに障害が発生すると、チャンネル制

50

御部 110 もしくはディスク制御部 140 から管理端末 160 に対してその旨の通知がなされる (S1311)。管理端末 160 は、障害が発生した旨の通知を受信すると、その旨のメッセージをディスプレイに表示したり警告音を出すなどして管理端末 160 のオペレータに知らせる。これにより管理端末 160 のオペレータ等は各チャンネル制御部 110 で実行されている OS701 やアプリケーションプログラムに障害が発生していることを知ることができる。

【0064】

また、あるチャンネル制御部 110 で実行されている OS701 やアプリケーションプログラムに障害が発生した場合にはその旨が内部 LAN151 を介して他のチャンネル制御部 110 で実行される NAS マネージャ 706 に対して通知される。この通知は、チャンネル制御部 110 やディスク制御部 140 から他のチャンネル制御部 110 で実行されている NAS マネージャ 706 に対し、内部 LAN151 を介して直接的に行われる場合 (S1312) と、管理端末 160 を介して間接的に行われる場合とがある (S1311, S1313)。このうちチャンネル制御部 110 から他のチャンネル制御部 110 に対して直接的に通知が行われる仕組みでは、例えば管理端末 160 における障害など他の機器の障害に影響されることがなく、障害が発生している箇所や障害の要因 (障害要因がチャンネル制御部 110 にあるのかそうではないのか) を特定し易くなる。また、他の装置を中継しない分、通知は迅速に行われることになる。一方、管理端末 160 を介してチャンネル制御部 110 で実行される NAS マネージャ 706 に対して間接的に通知する仕組みの場合には、その通知先となるチャンネル制御部 110 を管理端末 160 から指定することが可能である。このため、例えば通知する必要のないチャンネル制御部 110 に対して通知しないようにすることができる。また、他のチャンネル制御部 110 への通知を管理端末 160 に任せてしまうことで、各チャンネル制御部 110 の構成も簡素化されるととともに、チャンネル制御部 110 側の処理負荷も軽減される。さらに何らかの障害によりチャンネル制御部 110 から直接他のチャンネル制御部 110 に対して障害が発生した旨の通知を行えないような場合でも、管理端末 160 から他のチャンネル制御部 110 に対して通知が行われることで、他のチャンネル制御部 110 に対して確実に障害の発生を通知することができる。

【0065】

ところで、NAS マネージャ 706 は、外部 LAN400 に接続する情報処理装置 200 に対して Web サーバとして機能している。NAS マネージャ 706 は、チャンネル制御部 110 やディスク制御部 140 などから通知される OS701 やアプリケーションプログラムの障害の状況を記載した Web ページを生成し、この Web ページを外部 LAN400 を通じてアクセスしてくる情報処理装置 200 に対して提供する (S1314)。情報処理装置 200 のユーザやオペレータ等は、この Web ページを利用して各チャンネル制御部 110 で実行されている OS701 やアプリケーションプログラムの障害の発生の有無を知ることができる。

【0066】

また、本発明のストレージシステム 600 においては、管理端末 160 や NAS マネージャ 706 に対する各チャンネル制御部 110 で実行されている OS701 やアプリケーションプログラムにおける障害発生の通知はチャンネル制御部 110 やディスク制御部 140 が自発的 (プッシュ式) に行なう。すなわち、チャンネル制御部 110 やディスク制御部 140 はあるチャンネル制御部 110 で実行されている OS701 やアプリケーションプログラムに何らかの障害が発生したことを検知すると、その旨の通知を管理端末 160 や他のチャンネル制御部 110 に対して自発的に行う。

【0067】

このように本実施例におけるストレージシステム 600 においては、あるチャンネル制御部 110 において実行されている OS701 やアプリケーションプログラムにおいて障害が発生した旨の通知がチャンネル制御部 110 やディスク制御部 140 から自発的に管理端末 160 や他のチャンネル制御部 110 等に対してなされる。このためチャンネル制御部 110 で実行される OS701 やアプリケーションプログラムに障害が発生した場合には、その

旨の通知を管理端末160や他のチャンネル制御部110、情報処理装置200に対して迅速に行うことができる。従って、ユーザやオペレータ等は、各チャンネル制御部110において障害が発生していることを早期に知ることができ、フェイルオーバー処理や復旧処理を円滑に進めることが可能となる。またこのように通知が自発的になされることで、ポーリング処理などによる他の装置を監視するための仕組みが不要となり、管理端末160やチャンネル制御部110の構成が単純化され、またこれらの処理負荷も軽減されることになる。

【0068】

以下、以上に説明した本実施例のストレージシステム600におけるチャンネル制御部110で実行されているOS701やアプリケーションプログラムについての障害監視の仕組みについてより具体的に説明する。

【0069】

=== ダンプ情報 ===

まずチャンネル制御部110で実行されるOS701から出力される上述のダンプ情報について説明する。ダンプ情報とは、OS701で実行されるプロセスが異常終了した場合やOS701自身の異常などに起因して、OS701から出力される情報である。図14にダンプ情報のデータフォーマットの一例を示している。データフォーマット1400において、「NAS装置番号」の欄にはチャンネル制御部110のIDが設定される。「ネットワークアドレス」の欄にはチャンネル制御部110のネットワークアドレスが設定される。ネットワークアドレスとは、例えば内部LAN151上の装置に付与されるIPアドレスである。「生成日時」の欄にはそのダンプ情報が生成された日時が設定される。「ダンプデータ」の欄にはダンプ情報の実体が設定される。実体とは、例えばプロセスの異常終了時におけるメモリイメージである。なお、OS701が例えばUNIX（登録商標）である場合には、ダンプ情報はOS701で実行されるプロセスが特定のシグナルを受信したことを契機としてOS701から出力される。ダンプ情報は例えばプロセスやOS701の異常終了の原因究明に利用されたり、プログラムのデバッグなどに利用される。

【0070】

本実施例のストレージシステム600においては、各チャンネル制御部110で実行されているOS701から出力されるダンプ情報は、LUに記憶される。このLUはユーザデータが記憶されるLUとは異なるダンプ情報を記憶するために専用に分けられる。以下、ダンプ情報が格納されるLUのことを、ダンプ専用LUと称する。このようにダンプ情報を専用のLUに記憶するようにしているのは、例えばダンプ情報が記憶されることによりユーザ用の記憶領域が浸食されてしまうのを防ぐためである。なお、ダンプ専用LUを設けずにユーザデータと同じLUにダンプ情報を格納する構成とすることもできることはもちろんである。また、本実施例のストレージシステムにおいては、ダンプ専用LUは各チャンネル制御部110に対応させて分けられることもあるし、共用される場合すなわち1つのダンプ専用LUに異なるチャンネル制御部で実行されるOS701から出力されるダンプ情報が記憶されるように設定されることもある。このようにダンプ専用LUを共用されるように運用している場合には、共用されないように運用されている場合に比べ、ダンプ情報を格納するためのLUの数を減らすことができるという利点がある。

【0071】

ダンプ専用LUが複数のチャンネル制御部110に共用されるように運用している場合には、さらにそのダンプ専用LUの記憶領域を複数の記憶領域（以下、パーティションと記す）に区分して管理し、パーティションごとに1つのチャンネル制御部110のダンプ情報が記憶されるようにすることもできる。この場合には例えば図15に示す記憶領域管理テーブル1500のように、各LUにそのLUのパーティションに記憶されているダンプ情報に関する情報が管理される。なお、図15において、「格納域」の欄にはLU内に設定されているパーティションの識別子が設定される。「ダンプデータの有無」の欄にはダンプ情報が書き込まれているかどうかを示す情報が設定される。「NAS装置のID」の欄にはそのパーティションが割り当てられているチャンネル制御部110のIDが設定される。

「ダンプ生成日時」の欄にはダンプ情報が生成された日時が設定される。このようにパーティションごとにそのダンプ情報を記憶することができるチャンネル制御部110が設定されるようにすることで、例えばあるチャンネル制御部110から出力されるダンプ情報が他のチャンネル制御部110から出力されるダンプ情報により上書きされてしまうのを防ぐことができる。

【0072】

===チャンネル制御部から通知===

次に、あるチャンネル制御部110において実行されるOS701やアプリケーションプログラムに何らかの障害が発生した場合に、その旨の通知がそのチャンネル制御部110から管理端末160もしくは他のチャンネル制御部110に対して自発的になされる仕組みについて、図16に示すフローチャートとともに説明する。あるチャンネル制御部110で実行されているOS701が、当該OS701自身や当該OS701上で実行されているアプリケーションプログラムに何らかの障害が発生したことを検知する(S1611)。ここでチャンネル制御部110は、例えば、障害に関する割り込み処理の発生や、チャンネル制御部110で実行されるOS701がダンプ情報を出力したことを検知することにより障害が発生したことを認知する。次にチャンネル制御部110は、自身に障害が発生した旨を記載したメッセージを内部LAN151を通じて管理端末160もしくは他のチャンネル制御部110に対して送信する(S1612)。

10

【0073】

上記メッセージのデータフォーマットの一例を図17に示している。このデータフォーマット1700において「障害発生中のNAS識別子」の欄には、障害が発生しているチャンネル制御部110のIDが設定される。「ネットワークアドレス」の欄には、上記チャンネル制御部110に割り当てられている内部LAN151上のネットワークアドレスが設定される。ネットワークアドレスは例えばIPアドレスである。「障害発生日時」の欄には、その障害が発生した日時が設定される。「ダンプ情報記憶先LU」の欄には、そのチャンネル制御部110で実行されるOS701のダンプ情報の出力先のLUの識別子であるLUN(Logical Unit Number)が設定される。「格納域」の欄には、そのチャンネル制御部110の出力先として割り当てられているパーティションの識別子(ID)が設定される。「障害発生要因」の欄には障害の発生の原因がOS701の障害に起因するものであるのか、アプリケーションプログラムに起因するものであるのかを示す情報が設定される。なお、他のチャンネル制御部110へのメッセージの送信に際しては、他のチャンネル制御部110が正常に動作しているかどうかを管理端末160やSNMPによる問い合わせにより事前に確認するようにすることもできる。

20

30

【0074】

再び図16に戻って説明を進める。例えばOS701が当該OS701自身の障害を検知した場合、OS701はLUにダンプ情報を出力する(S1613)。一方、管理端末160もしくは他のチャンネル制御部110のNASマネージャ706は、上記メッセージを受信すると(S1614)これをメモリ113に記憶する(S1615)。このようにしてメモリ113に記憶されたメッセージの内容は例えば管理端末160のディスプレイに表示される。また、上記メッセージの内容は、NASマネージャ706が提供するWeb

40

【0075】

管理端末160もしくはNASマネージャ706は、上記メッセージを送信してきたチャンネル制御部110からのダンプ情報の出力に関する処理が完了したかどうかを内部LAN151を通じて監視している(S1616)。そして、管理端末160もしくはNASマネージャ706は、ダンプ情報の出力が完了したことを認知すると、上記メッセージを送信してきたチャンネル制御部110に対してリセット信号を送信する(S1617)。リセット信号を受信したチャンネル制御部110は(S1618)自身を再起動する(S161

50

9) ように制御する。このように管理端末 160 や N A S マネージャ 706 により障害が発生しているチャンネル制御部 110 のリセット処理が自動的になされることで、障害の発生から復旧までにかかる時間が短くて済み、ユーザに対する影響を極力低減することができる。

【0076】

=== ディスク制御部から通知 ===

あるチャンネル制御部 110 において何らかの障害が発生した場合には、ディスク制御部 140 からその旨の通知が管理端末 160 もしくは他のチャンネル制御部 110 に対してなされる。この場合の仕組みについて図 18 に示すフローチャートとともに具体的に説明する。

10

【0077】

ディスク制御部 140 は、記憶デバイス 300 に対して何らかのデータの書き込みが行われていることを検知すると (S1811)、その書き込み先となる L U がダンプ専用 L U であるかどうかを判断する (S1812)。この判断はデータの書き込み先に設定されている L U とディスク制御部 140 の N V R A M 144 に記憶されている、後述するダンプ専用 L U 管理テーブル 2000 (図 20) とを対照することにより行われる。

【0078】

上記データの書き込み先は、キャッシュメモリ 130 に例えば、図 19 に示すデータフォーマットで記憶されている。このデータフォーマット 1900 には、この書き込みデータをキャッシュメモリ 130 に設定したチャンネル制御部 110 (つまりそのデータ書き込みの指示元であるチャンネル制御部 110) の ID、データの書き込み先となる L U の識別子 (L U N)、書き込み対象となるデータや書き込み先となる L U 上の記憶位置を特定する情報 (例えば先頭アドレス) などが設定されている。図 20 に示すように、ダンプ専用 L U 管理テーブル 2000 には、L U N ごとにそれがダンプ専用 L U であるのかそうでないのかが示されている。ディスク制御部 140 は、上記データフォーマットに設定されている L U N をダンプ専用 L U 管理テーブル 2000 に対照することにより書き込み先がダンプ専用 L U であるかどうかを判断する。そしてディスク制御部 140 は、上記判断でダンプ専用 L U が書き込み先であると判断すると (S1812: Y E S)、ダンプ専用 L U にデータを書き込む (すなわち、ダンプ情報をダンプ専用 L U に書き込む) (S1813) とともに、チャンネル制御部 110 に何らかの障害が発生した旨を通知するためのメッセージを管理端末 160 もしくは他のチャンネル制御部 110 に対して送信する (S1814)。

なお、他のチャンネル制御部 110 に対するメッセージの送信に際しては、他のチャンネル制御部 110 が正常に動作しているかどうかを管理端末 160 に問い合わせることにより事前に確認することができる。なお、ここで送信されるメッセージの内容は例えば図 17 に示したものと同様である。

20

30

【0079】

管理端末 160 もしくは他のチャンネル制御部 110 は、メッセージを受信する (S1815) とその内容を自身のメモリに記憶する (S1816)。一方、(S1812: N O) において前記データ書き込み要求がダンプ専用 L U 以外の L U を対象とするものである場合には、ディスク制御部 140 はデータの書き込みは行う (S1817) が、上記メッセージの送信は行わない。以上に説明した仕組みにより、ディスク制御部 140 から管理端末 160 もしくは他のチャンネル制御部 110 に対し、チャンネル制御部 110 において障害が発生している旨の自発的な通知が行われることになる。

40

【0080】

ところで、このように O S 701 からダンプ情報が出力された旨の通知がディスク制御部 140 の仕組みによりディスク制御部 140 から自発的に行われることで、チャンネル制御部 110 と、管理端末 160 もしくは他のチャンネル制御部 110 との間の通信に何らかの障害が発生して管理端末 160 に障害発生のお知らせを行えないような場合でも、O S 701 からダンプ情報が出力された旨の通知を管理端末 160 や他のチャンネル制御部 110 に対して通知することができる。なお、このような場合としては、例えば、O S 701 の障害

50

により、OS701がダンプ情報を書き込むことはできるが、管理端末160もしくは他のチャンネル制御部110に対して障害が発生した旨の通知を行えない場合があげられる。

【0081】

また、ディスク制御部140と管理端末160もしくは他のチャンネル制御部110との間の通信に何らかの障害が発生している場合には、上述した仕組みによりチャンネル制御部110から管理端末160もしくは他のチャンネル制御部110に対して障害が発生している旨の通知が自発的になされる。このように本発明のストレージシステム600においては、チャンネル制御部110からの通知とディスク制御部140からの通知とが相補的に行われることで、あるチャンネル制御部110において障害が発生している旨の通知が、管理端末160や他のチャンネル制御部110に対して確実に行なわれることになる。さらにこのようにチャンネル制御部110とディスク制御部140の双方から相補的に通知が行われる場合には、各経路からの通知が正常に行われているかどうかを確認することで、ストレージシステム600の構成要素のどの部分に障害が発生しているのかを推測することも可能となる。

10

【0082】

=== 管理端末からの通知 ===

ところで、以上に説明した障害管理の仕組みにおいては、あるチャンネル制御部110において障害が発生した場合における他のチャンネル制御部110へのその旨の通知はチャンネル制御部110もしくはディスク制御部140から直接他のチャンネル制御部110に対して行われている。しかしながらこの通知は、チャンネル制御部110において障害が発生した旨の通知を受信した管理端末160を介して間接的に他のチャンネル制御部110に対して行うようにすることができる。図21に示すフローチャートとともにこの仕組みについて説明する。

20

【0083】

管理端末160は、チャンネル制御部110もしくはディスク制御部140から送信されてくる(S2111)障害が発生した旨のメッセージを受信すると(S2112)、そのメッセージの内容を記憶する(S2113)。次に管理端末160は、そのメッセージに送信元として設定されているチャンネル制御部110とは異なる他のチャンネル制御部110(すなわち他のチャンネル制御部110)に対し、そのメッセージの送信元であるチャンネル制御部110において障害が発生している旨を設定したメッセージを送信する(S2114)。ここで送信先となるチャンネル制御部110は、管理端末160が記憶している、図22に一例を示す通知先テーブル2200に登録されているチャンネル制御部110が対象となる。図22では「通知先」の欄には通知先となるチャンネル制御部のネットワークアドレス(IPアドレス)が、また「装置種類」の欄に通知先となる装置の種類が設定されている。通知先テーブル2200の内容は、例えば管理端末160を操作するオペレータ等により設定される。

30

【0084】

他のチャンネル制御部110は、このようにして上記メッセージが送られてくるとこれを受信する(S2115)。そしてこのメッセージの内容は、例えばNASマネージャ706に利用される。なお、受信したメッセージの内容はチャンネル制御部110で実行されているNASマネージャ706に通知される。

40

【0085】

このように他のチャンネル制御部110への通知が管理端末160を介して間接的になされることで、例えば、通知先となるチャンネル制御部110を管理端末160から特定することが可能となる。すなわち障害が発生している旨の通知をする必要のないチャンネル制御部110には通知を行わないようにすることができる。また、同じ通知が複数のチャンネル制御部110に重複して行われるのを防ぐことができる。また、他のチャンネル制御部110への通知が管理端末160を介して間接的になされることで、各チャンネル制御部110の構成が簡素化されチャンネル制御部110側の処理負荷が軽減される。さらに、何らかの障害によりチャンネル制御部110から他のチャンネル制御部110に障害が発生した旨の通知

50

を直接行えないような場合でも、管理端末 160 から他のチャンネル制御部 110 に通知を行うことが可能となる。

【0086】

=== 管理端末の機能 ===

管理端末 160 は、各チャンネル制御部 110 で実行される OS 701 やアプリケーションプログラムの実行状態を監視するためのユーザインタフェースを備えている。管理端末 160 は、チャンネル制御部 110 やディスク制御部 140 から送信されてくる障害が発生した旨の通知を受信して、その通知の内容をオペレータ等に知らせる画面（以下、障害通知画面と記す）をディスプレイに表示する。

【0087】

図 23 に障害通知画面の一例を示している。この図に示す障害通知画面 2300 において、「チャンネル制御部 ID」の欄には各チャンネル制御部 110 に固有に割り当てられている識別子が表示される。「状態」の欄には各チャンネル制御部 110 の動作状態が表示され、正常に動作している場合には『正常』が、障害（異常）が発生している場合には『障害』が表示される。「通知元」の欄には、「状態」の欄に『障害』が表示されているチャンネル制御部 110 についてその障害通知がどこから通知されたのかが示されており、その通知を行った装置の内部 LAN 151 上のネットワークアドレス（例えば IP アドレス）が表示される。オペレータ等は、障害通知画面 2300 を参照することにより各チャンネル制御部 110 における障害発生の有無を知ることができる。またオペレータ等は「通知元」の欄を参照することにより障害の原因を推測できる。例えば「通知元」の欄にチャンネル制御部 110 に割り当てられているネットワークアドレスが表示されている場合には、そのチャンネル制御部 110 自身が障害の発生を通知しているので、少なくとも通知がなされた時点ではそのチャンネル制御部 110 の通知機能は正常に動作していたものと推測することができる。また「通知元」の欄にディスク制御部 140 に割り当てられているネットワークアドレスが表示されている場合には、チャンネル制御部 110 の通知機能に何らかの障害が発生しているものと推測することができる。

【0088】

=== 障害の統合的な管理 ===

あるチャンネル制御部 110 で実行される OS 701 やアプリケーションプログラムに何らかの障害が発生した場合には、その旨がその障害を検知した OS 701 やアプリケーションプログラムから NAS マネージャ 703 に通知される。またチャンネル制御部 110 は障害が発生した旨を他のチャンネル制御部 110 で実行されている NAS マネージャ 706 に通知する。また障害管理プログラム 705 は、クラスタを構成している他のチャンネル制御部 110 における障害の発生を検知すると、その旨を NAS マネージャ 706 に対して通知する。このように各チャンネル制御部 110 で実行されている NAS マネージャ 706 には、ストレージシステム 600 の全体における障害に関する情報が集約され、これによりストレージシステム 600 における障害の統合的な管理が可能となる。

【0089】

NAS マネージャ 706 は、障害発生のお知らせを受信すると、その旨やその通知に基づいて生成したチャンネル制御部 110 の障害に関する情報を、電子メール、SNMP、TCP/IP 等により情報処理装置 200 や管理端末 160 に対して通知する。図 24 は電子メール等により通知されるデータの内容の一例である。この図において、「障害発生中の NAS 識別子」の欄には障害が発生しているチャンネル制御部 110 の ID が設定される。「ネットワークアドレス」の欄には障害が発生しているチャンネル制御部 110 のネットワークアドレス（例えば IP アドレス）が設定される。「障害発生日時」の欄にはチャンネル制御部 110 において障害が発生した日時が設定される。「ダンプ情報記憶先 LU」の欄には障害が発生したチャンネル制御部 110 から出力されたダンプ情報が格納されている LU の識別子である LUN が設定される。「格納域」の欄にはダンプ情報が格納されるパーティションの ID が設定される。このように障害に関する情報が電子メールにより情報処理装置 200 に通知されるため、情報処理装置 200 のユーザやオペレータ等は障害が発生し

10

20

30

40

50

ているチャンネル制御部 110 を容易に特定することが可能となる。また障害が発生した場合に原因究明や障害復旧のための処理を円滑に行うことができる。

【0090】

またNASマネージャ706は、他のチャンネル制御部110や管理端末160から通知されるメッセージに基づいてストレージシステム600の各チャンネル制御部110についての障害の状況を記載したWebページを生成する仕組みを備えている。図25はそのようなWebページの一例である。このWebページ2500において、「障害NAS装置名称」の欄には障害が発生しているチャンネル制御部110の名称が設定される。「障害発生日時」の欄には障害が発生した日時が設定される。「障害内容」にはその障害がOS701によるものであるのか、アプリケーションによるものであるのかといった障害の状況を把握するために必要な情報が設定される。情報処理装置200からこのWebページにアクセスすることで、情報処理装置200のユーザやオペレータ等はストレージシステム600のチャンネル制御部の障害の状況を知ることができる。

10

【0091】

またNASマネージャ706は、障害の発生の通知があると、例えば障害管理プログラム705と連携してチャンネル制御部110間でのフェイルオーバや電子メール等による通知等の処理を実行する。この仕組みを図26に示すフローチャートとともに説明する。

【0092】

障害管理プログラム705は、OS701やアプリケーションプログラムの障害発生の有無を設定されたタイミングで監視している(S2611)。障害管理プログラム705は障害の発生を検知すると(S2612: YES)、障害の発生をNASマネージャ706に通知する(S2613)。一方、OS701やアプリケーションプログラムも、障害の発生を検知すると(S2614)、その旨をNASマネージャ706に通知する(S2615)。また、管理端末160も上述した仕組みによりあるチャンネル制御部110において動作するOS701に障害が発生した旨を検知すると(S2616)、その旨をNASマネージャ706に通知する(S2617)。そして、NASマネージャ706は、障害管理プログラム705、OS701やアプリケーションプログラム、管理端末160、のそれぞれから送られてくる障害発生の通知を受信すると(S2618)、フェイルオーバ、電子メール等による情報処理装置200への通知、障害に関する情報を記載したログ情報のLUへの出力などを行う。このようにNASマネージャ706を中心として進められる統合的な障害管理の仕組みにより、障害管理が一元化され、システムの運用・管理が容易となる。

20

30

【0093】

=== 障害発生箇所の特定方法 ===

ところで、電子メール等により情報処理装置200に対して何らかの通知を行うように、管理端末160から各チャンネル制御部110で実行されるNASマネージャ706に対して指示した場合、その通知機能に障害のあるチャンネル制御部110からはその通知が行われず、これにより障害が発生しているチャンネル制御部110を特定することができる。図27とともにこの仕組みについて具体的に説明する。

【0094】

チャンネル制御部110において障害が発生すると(S2711)その旨の通知が管理端末160になされる(S2712)。管理端末160はこの通知を受信すると(S2713)、そのチャンネル制御部110を含む全てのチャンネル制御部110に対してその旨の通知を電子メール等で情報処理装置200に通知するように指示する(S2714)。この指示を受信した(S2715)各チャンネル制御部110は、電子メール等により上記障害の発生についての通知を行う(S2716)。そして、情報処理装置200はその電子メール等による通知を受信する(S2717)。ここで電子メール等の通知が行えないチャンネル制御部110からは電子メール等による通知が情報処理装置200に対してなされず、これにより情報処理装置200のオペレータ等は通知がこないチャンネル制御部110に障害が発生しているものと推定することができる(S2718)。

40

50

【 0 0 9 5 】

＝ ＝ ダンプ情報通知機能 ＝ ＝

各チャンネル制御部 1 1 0 で実行される N A S マネージャ 7 0 6 は、ダンプ情報の内容を電子メールに記載して情報処理装置 2 0 0 に送信する機能を備えている。ここでは N A S マネージャ 7 0 6 からダンプ情報の内容が付帯された電子メールが情報処理装置 2 0 0 に送信される場合を例として図 2 8 に示すフローチャートとともにその仕組みについて説明する。

【 0 0 9 6 】

あるチャンネル制御部 1 1 0 に何らかの障害が発生すると、そのチャンネル制御部 1 1 0 の O S 7 0 1 がダンプ専用 L U にダンプ情報を出力する。また障害が発生した旨が記載されたメッセージがそのチャンネル制御部 1 1 0 またはディスク制御部 1 4 0、もしくは管理端末 1 6 0 から他のチャンネル制御部 1 1 0 に送信される (S 2 8 1 1)。ここで上記メッセージのデータフォーマットは、例えば図 1 7 に示したものと同一である。他のチャンネル制御部 1 1 0 で実行される N A S マネージャ 7 0 6 は、上記メッセージを受信すると (S 2 8 1 2) 上記ダンプ情報が記憶されている L U にアクセス可能となるようにストレージシステム 6 0 0 のアクセス制限の設定を変更した後 (S 2 8 1 3)、上記メッセージから特定される L U から内部 L A N 1 5 1 を通じてダンプ情報を読み出す (S 2 8 1 4)。なお、L U に記憶領域が区分されている場合には、上記メッセージに設定されている「格納域」の欄の内容を参照することにより特定される。

【 0 0 9 7 】

次に他のチャンネル制御部 1 1 0 で実行されている N A S マネージャ 7 0 6 は、L U から読み出したダンプ情報からキャッシュ情報などのユーザ等に参照させる必要のない情報の除去やデータ圧縮などの電子メールに付帯させるための参照用のダンプ情報を生成する (S 2 8 1 5)。また、ダンプ情報はそのままではファイルシステムが利用することができない形式で L U に記憶されていることがあるが、その場合にはファイルシステムから参照可能な形式 (例えばテキストデータ形式) にダンプ情報を変換する処理も行われる。次に他のチャンネル制御部 1 1 0 で実行される N A S マネージャ 7 0 6 は、生成された参照用のダンプ情報を付帯させた電子メールを生成し、生成した電子メールを情報処理装置 2 0 0 に送信する (S 2 8 1 6)。情報処理装置 2 0 0 はその電子メールを受信する (S 2 8 1 7)。なお、他のチャンネル制御部 1 1 0 で実行される上記 N A S マネージャ 7 0 3 は、電子

【 0 0 9 8 】

このように障害が発生するとそのダンプ情報が電子メールにより情報処理装置 2 0 0 のユーザやオペレータ等に迅速に通知される。従って、ユーザやオペレータ等は障害原因の究明や障害復旧のための対策等の作業を円滑に進めることができる。また、ユーザやオペレータ等は、チャンネル制御部 1 1 0 において発生した障害の原因究明や障害復旧を行うためにダンプ情報の所在を特定する作業や特有の形式で L U に記憶されているダンプ情報を参照可能な形式に変換するといったダンプ情報の収集のための面倒な作業から開放され、障害原因の究明や障害復旧のための対策等の作業を迅速に進めることができる。

【 0 0 9 9 】

なお、以上では N A S マネージャ 7 0 6 が主体となって情報処理装置 2 0 0 にダンプ情報を提供する構成について説明したが、管理端末 1 6 0 と情報処理装置 2 0 0 とを通信可能に接続することで管理端末 1 6 0 が主体となって情報処理装置 2 0 0 に通知する構成とすることもできる。またチャンネル制御部 1 1 0 において障害が発生した旨の通知を受信した管理端末 1 6 0 が他のチャンネル制御部 1 1 0 で実行される N A S マネージャ 7 0 6 を制御してダンプ情報を記載した電子メールを送信するようにしてもよい。なお、この場合には N A S マネージャ 7 0 6 が参照用のダンプ情報を生成する構成としてもよいし、管理端末 1 6 0 が参照用のダンプ情報を生成してこれを N A S マネージャ 7 0 6 に提供する構成としてもよい。また、上記の説明は N A S マネージャ 7 0 6 がダンプ情報を電子メールに添付する構成としているが、電子メール以外にも S N M P や T C P / I P 等の通信によりダ

10

20

30

40

50

ンプ情報を情報処理装置 200 に提供する構成とすることもできる。

【0100】

以上本実施の形態について説明したが、上記実施例は本発明の理解を容易にするためのものであり、本発明を限定して解釈するためのものではない。本発明はその趣旨を逸脱することなく変更、改良され得ると共に、本発明にはその等価物も含まれる。

【0101】

【発明の効果】

ストレージシステムの制御方法、ストレージシステム、及びプログラムを提供することができる。

【図面の簡単な説明】

【図1】本実施の形態に係るストレージシステムの全体構成を示すブロック図である。

【図2】本実施の形態に係る管理端末の構成を示すブロック図である。

【図3】本実施の形態に係る物理ディスク管理テーブルを示す図である。

【図4】本実施の形態に係るLU管理テーブルを示す図である。

【図5】本実施の形態に係るストレージシステムの外観構成を示す図である。

【図6】本実施の形態に係る記憶デバイス制御装置の外観構成を示す図である。

【図7】本実施の形態に係るチャンネル制御部のハードウェア構成を示す図である。

【図8】本実施の形態に係るメモリに記憶されるデータの内容を説明するための図である。

【図9】本実施の形態に係るメタデータの一例を示す図である。

【図10】本実施の形態に係るディスク制御部のハードウェア構成を示す図である。

【図11】本実施の形態に係るストレージシステムのソフトウェア構成を示す図である。

【図12】本実施の形態に係る排他制御及びフェイルオーバ制御を説明するためのブロック図である。

【図13】本実施の形態に係る、障害管理の仕組みを説明する図である。

【図14】本実施の形態に係る、ダンプ情報のデータフォーマットの一例を示す図である。

【図15】本実施の形態に係る、記憶領域管理テーブルの一例を示す図である。

【図16】本実施の形態に係る、チャンネル制御部において障害が発生した旨の通知が管理端末もしくは他のチャンネル制御部に対して自発的になされる仕組みを説明するフローチャートである。

【図17】本実施の形態に係る、障害が発生した旨を記載したメッセージのデータフォーマットの一例を示す図である。

【図18】本実施の形態に係る、チャンネル制御部において障害が発生した旨の通知がディスク制御部から管理端末もしくは他のチャンネル制御部に対して自発的になされる仕組みを説明するフローチャートである。

【図19】本実施の形態に係る、キャッシュメモリに記憶されているデータの書き込み先を示すデータのフォーマットを示す図である。

【図20】本実施の形態に係る、ダンプ専用LU管理テーブルの一例を示す図である。

【図21】本実施の形態に係る、チャンネル制御部において障害が発生した旨の通知を管理端末から間接的に他のチャンネル制御部に対して行う仕組みを説明するフローチャートである。

【図22】本実施の形態に係る、通知先テーブルの一例を示す図である。

【図23】本実施の形態に係る、障害通知画面の一例を示す図である。

【図24】本実施の形態に係る、NASマネージャから情報処理装置や管理端末に対して電子メール等により通知されるデータの内容の一例を示す図である。

【図25】本実施の形態に係る、NASマネージャが生成するWebページの一例を示す図である。

【図26】本実施の形態に係る、NASマネージャに障害の発生した旨の通知があった場合にフェイルオーバや電子メール等による通知などの処理が実行される仕組みを説明するフロー

10

20

30

40

50

チャートである。

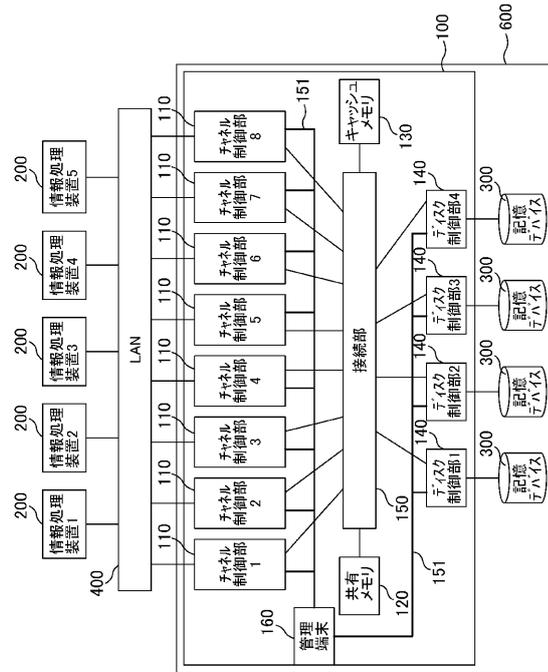
【図 27】本実施の形態に係る、障害の発生の通知を受信した N A S マネージャが必ずその旨を電子メール等により情報処理装置に通知するようにすることにより障害が発生しているチャンネル制御部を推測する仕組みを説明するフローチャートである。

【図 28】本実施の形態に係る、ダンプ情報の内容を電子メールに記載して情報処理装置に送信する仕組みを説明するフローチャートである。

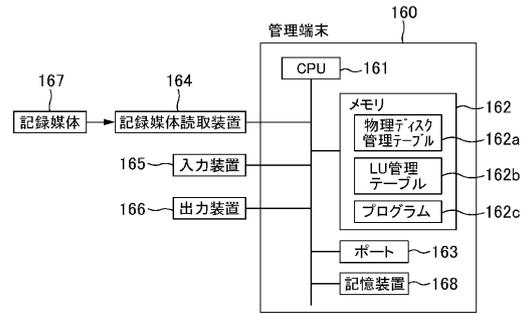
【符号の説明】

1 0 0	記憶デバイス制御装置	
1 1 0	チャンネル制御部 (C H N)	
1 1 1	ネットワークインタフェース部	10
1 1 2	C P U	
1 1 3	メモリ	
1 1 4	入出力制御部	
1 1 5	N V R A M	
1 1 6	ボード接続用コネクタ	
1 1 7	通信コネクタ	
1 1 8	回路基板	
1 1 9	I / O プロセッサ	
1 2 0	共有メモリ	
1 3 0	キャッシュメモリ	20
1 4 0	ディスク制御部	
1 5 0	接続部	
1 5 1	内部 L A N	
1 6 0	管理端末	
2 0 0	情報処理装置	
3 0 0	記憶デバイス	
4 0 0	L A N	
6 0 0	ストレージシステム	
7 0 1	O S	
7 3 0	メタデータ	30

【図1】



【図2】



【図3】

162a 物理ディスク管理テーブル

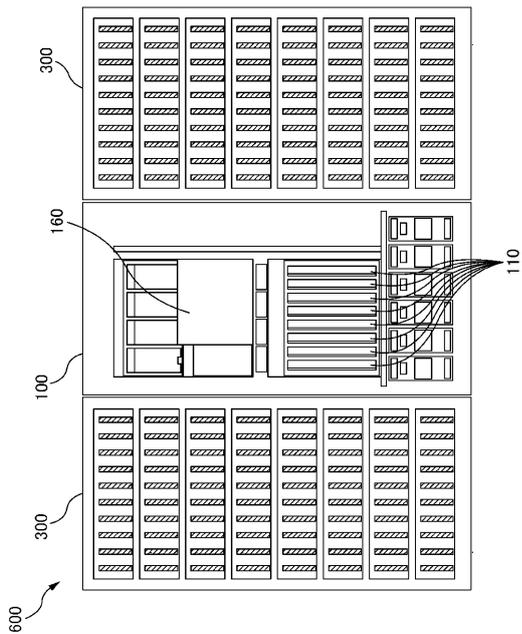
ディスク番号	容量	RAID	使用状況
#001	100GB	5	使用中
#002	100GB	5	使用中
#003	100GB	5	使用中
#004	100GB	5	使用中
#005	100GB	5	使用中
#006	50GB	—	未使用
...

【図4】

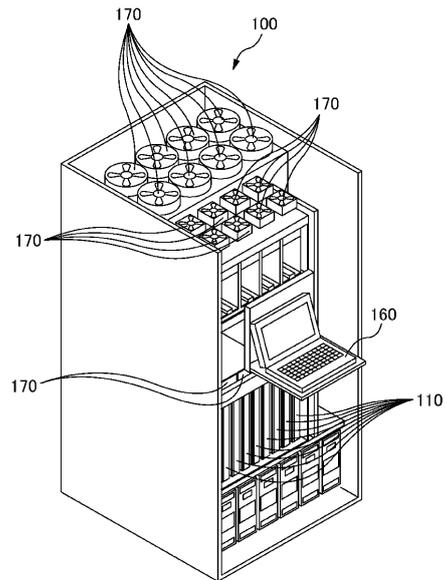
162b LU管理テーブル

LU番号	物理ディスク	容量	RAID
#1	#001,#002,#003,#004,#005	100GB	5
#2	#001,#002,#003,#004,#005	300GB	5
#3	#006,#007	200GB	1
...

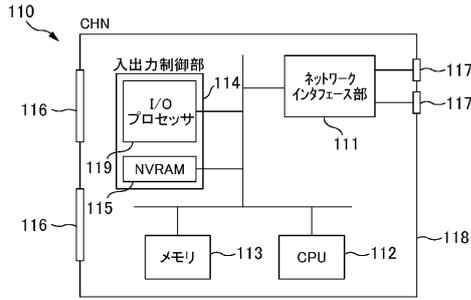
【図5】



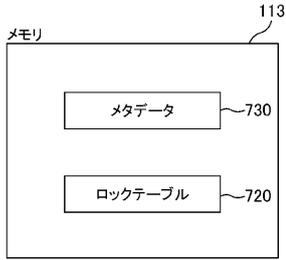
【図6】



【図7】



【図8】

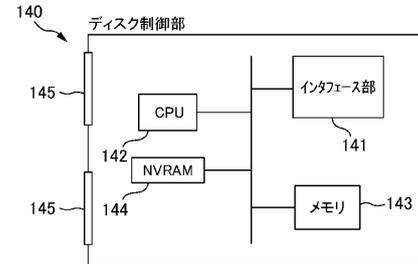


【図9】

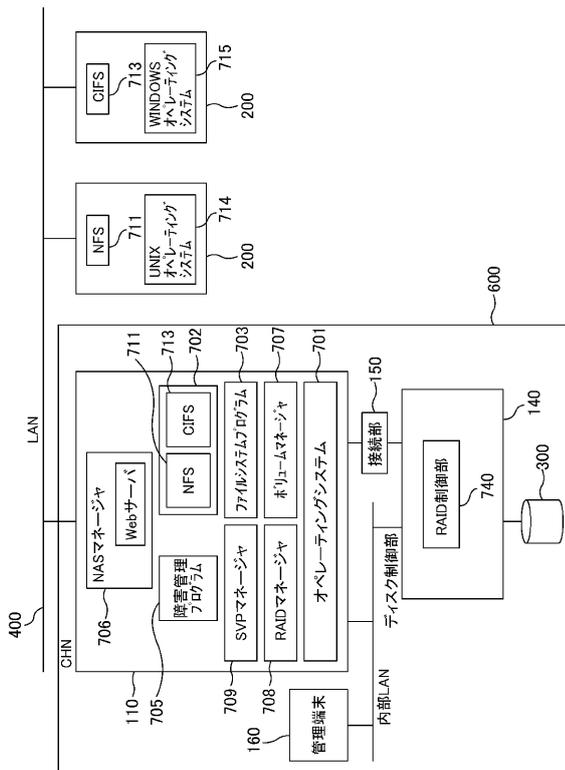
メタデータ

ファイル名	先頭アドレス	容量	所有者	更新時刻
A	7BSA	200MB	X	0:00
B	05BF	50MB	X	7:57
C	1F30	100MB	Y	9:15
D	470B	100MB	Z	15:20
⋮	⋮	⋮	⋮	⋮

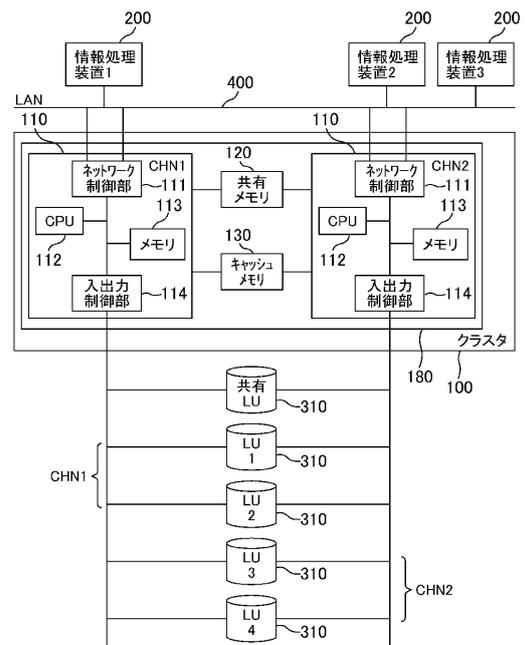
【図10】



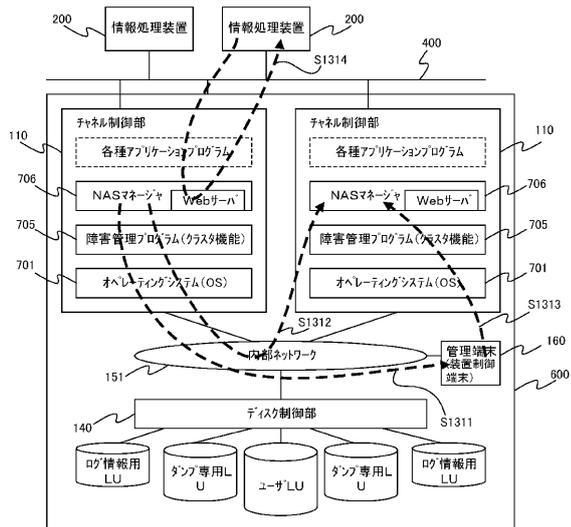
【図11】



【図12】



【図13】



【図15】

記憶領域(パーティション)管理テーブル

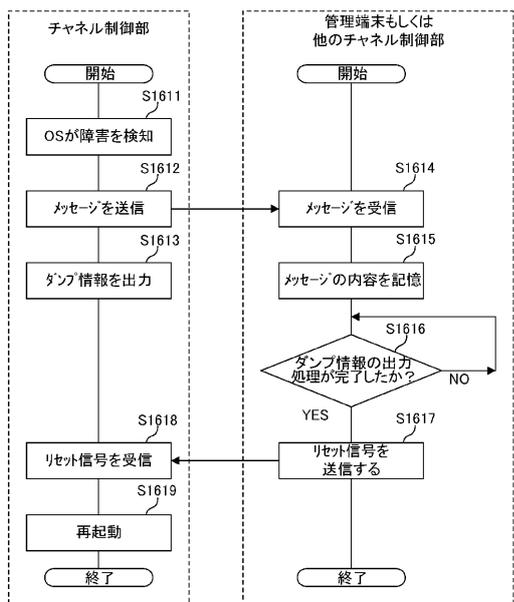
格納域(パーティションID)	ダンプ情報の有無	NAS装置のID	ダンプ生成日時
1	有	5	2003/01/11 03:45:11
2	無	-	-
3	有	1	2003/01/5 06:35:52
4	無	-	-

【図14】

ダンプ情報のデータ構造

項目	開始アドレス(byte)	内容の具体例
NAS装置番号	0	1
ネットワークアドレス	32	192.168.0.10
ダンプ生成日時	8	2003/01/11 03:45:11
ダンプデータ	128	0F15A0022811...

【図16】

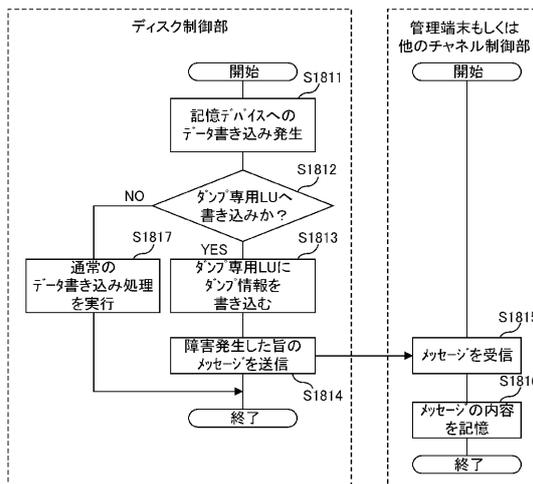


【図17】

メッセージのデータフォーマット

項目	内容(例)
障害発生中のNAS識別子	1
ネットワークアドレス	192.168.0.10
障害発生日時	2003/01/11 03:45:11
ダンプ情報記憶先LU	#1
格納域(パーティションID)	1
障害発生要因	OS

【図18】



【図 19】

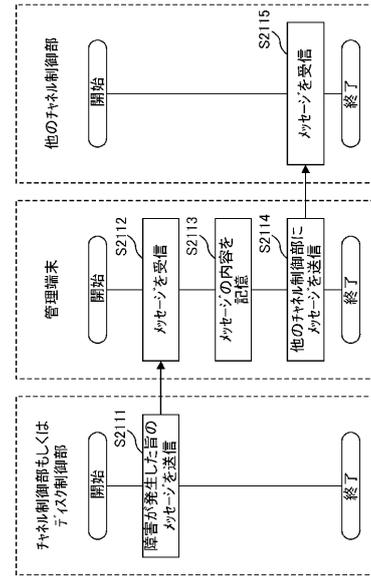
チャンネル制御部のID	書き込み先のLUN	書き込み先のアドレス
-------------	-----------	------------

【図 20】

ダンプ専用LU管理テーブル

LU番号	0 通常	1 ダンプ専用
#1	0	
#2	0	
#3	1	
⋮	⋮	⋮
⋮	⋮	⋮

【図 21】



【図 22】

通知先テーブル

通知先(IPアドレス)	装置種類
192.168.2.1	管理端末
192.168.2.11	チャンネル制御部
192.168.2.12	チャンネル制御部
192.168.2.13	チャンネル制御部
⋮	⋮
⋮	⋮

【図 24】

電子メール等により通知されるデータの例

項目	内容(例)
障害発生中のNAS識別子	1
ネットワークアドレス	192.168.0.10
障害発生日時	2003/01/11 03:45:11
ダンプ情報記憶先LU	LU1
格納域(パーティションID)	1

【図 23】

障害通知画面

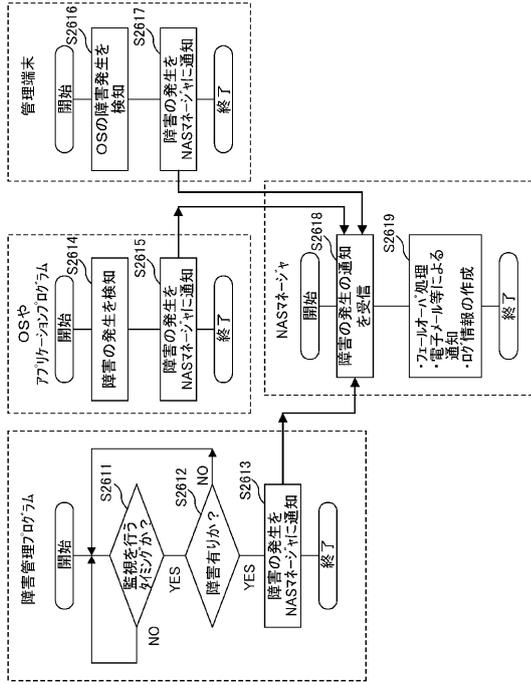
チャンネル制御部のID	状態	通知元
1	障害	192.168.0.10
2	正常	
3	正常	
4	正常	
5	正常	
6	正常	
7	正常	
8	正常	

【図 25】

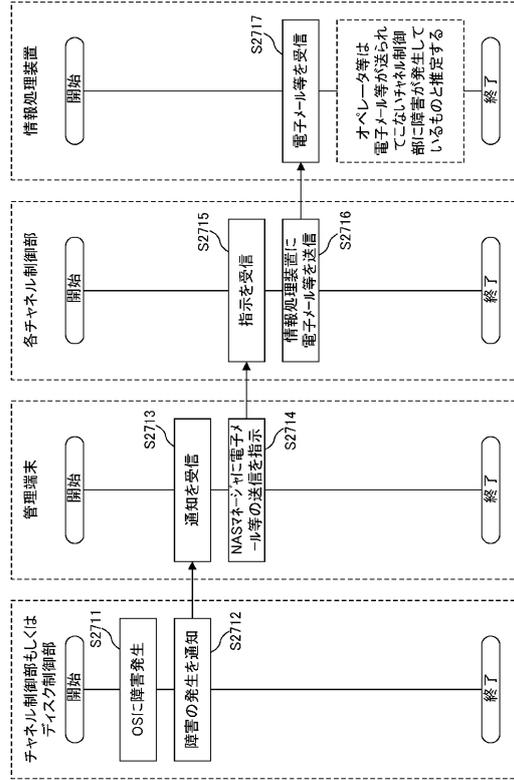
障害の内容を記載したWebページ

障害NAS装置名称	障害発生日時	障害内容
NAS装置1	2002/11/30 19:30:04	NAS OS障害
NAS装置4	2002/11/28 00:14:48	NAS OS障害

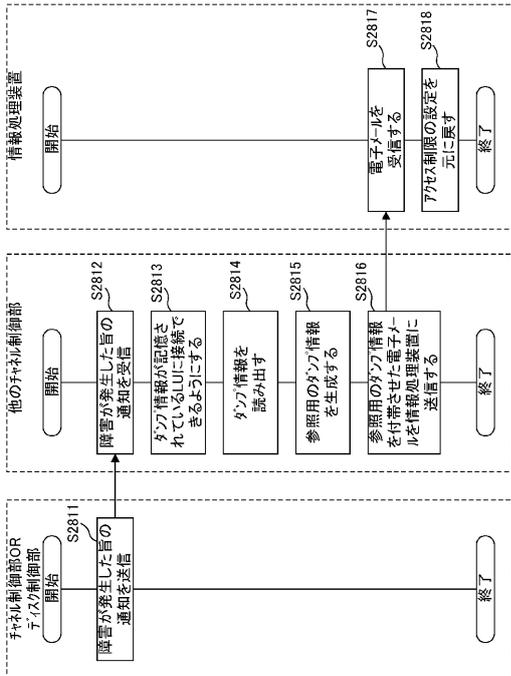
【 図 2 6 】



【 図 2 7 】



【 図 2 8 】



フロントページの続き

(72)発明者 藤本 修二

神奈川県横浜市戸塚区戸塚町5030番地 株式会社日立製作所ソフトウェア事業部内

Fターム(参考) 5B083 AA01 BB01 BB02 CE01 DD13 EE04 GG04